WHITEPAPER AND FINAL PERFORMANCE REPORT: "RECONSTRUCTING THE FIRST HUMANITIES COMPUTING CENTER"

NEH Digital Humanities Advancement Grant HAA-255991-17 USF, TAMPA, FL, 2017-2019

Steven E. Jones (PI), Howard Kaplan, Julianne Nyhan, Marco Passarotti, Geoffrey Rockwell, Paolo Senna, Stéfan Sinclair, Melissa Terras

BACKGROUND

In 1956, the Jesuit scholar Roberto Busa, S.J. founded the first dedicated center for humanities computing. (This is not the same as undertaking the first act of humanities computing. Nor does this historical "first"--however significant--mark the only intellectual lineage of what we now call the digital humanities.) Busa named the center CAAL, the Centro per L'Automazione dell'Analisi Letteraria--the Center for the Automation of Literary Analysis ("Letteraria" was sometimes replaced with "Linguistica"--or "Linguistic"). The idea was to use IBM punched-card machines to process language instead of numbers, to treat language as data. After several years in various temporary locations, in 1961 Busa's operation moved into a former textile factory in Gallarate, Italy, outside Milan, where IBM punched-card data processing machines were installed. Teams of student operators, most of them young women, in conjunction with scholars based at Father Busa's college in Gallarate, the Aloisianum, worked there for six years (1961-1967) on a lemmatized concordance to the works of St. Thomas Aquinas, the *Index Thomisticus*, as well as other projects in punched-card data processing. In the 1970s, the *Index Thomisticus* was finally printed in 56 bound volumes.

In other configurations and locations, in Boulder Colorado, for example, and back in Italy, CAAL continued to operate for years. But that six-year period in the former factory stands as a significant moment in the history of humanities computing, though its details have become obscured with time. The machinery was returned to IBM in 1967 and the main building was demolished sometime between January 2010 and March 2012. As evidence of the site, how it was actually configured and equipped, we are left with (1) Busa's own published accounts of the story of CAAL; (2) accounts by some of the former punched card operators; (3) the documentary record in the Busa Archive, including about 80 photographs, commissioned and staged by Busa himself, often as illustrations to press reports; (4) a few outside contemporary reports, for example in the typed manuscript still in the Busa Archive for a Dutch book on early humanities

computing by Félicien de Tollenaere, *Nieuwe wegen in de lexicologie* (1963); and (5) some documents in the IBM Archives and other collections.

PROJECT GOALS

This project is part of an ongoing widely-collaborative scholarly effort to "reconstruct" (by researching and modeling) the 1961-1967 instantiation of CAAL in its multiple dimensions, not in any final way or to settle every question, but in order to raise new questions and to reveal what we don't know about the first humanities computing center and the work that was done there. Our goal is to explore in detail the infrastructure, workflow, and the historical and institutional contexts for this significant site (literally and figuratively) in the history of humanities computing.

Busa's work was a historic milestone for humanities computing, one strand of which fed into what we now call the digital humanities, an interdisciplinary field of increasing significance in the academy and the wider public arena. Busa's work is often treated as a myth of the origin of humanities computing. We want to complicate the myth with history. One of our premises is that technology includes material and social infrastructure, and we've tried to illuminate issues important to the humanities broadly considered, such as the gendered labor involved in this kind of work, the specific roles of Cold-War funding and corporate sponsorship for humanities research, and the emergence of new forms of interdisciplinary scholarship across science and the humanities.

PROJECT TEAM AND CONTRIBUTORS

The **Project Team** consisted of the following personnel:

Steven E. Jones, University of South Florida (Project Director)

Howard Kaplan, University of South Florida, Advanced Technologies Manager, Advanced Visualization Center (AVC) (Expert in charge of 3D development)

Julianne Nyhan, University College London (Expert Adviser)

Marco Carlo Passarotti, Director of CIRCSE, Università Cattolica del Sacro Cuore, Milan, Italy (computational linguist and Busa Archive coordinator)

Geoffrey Rockwell, University of Alberta, Canada (Expert Adviser)

Paolo Senna, Librarian and Archivist, Università Cattolica del Sacro Cuore, Milan, Italy (Expert Adviser, Busa Archive coordinator)

> **Stéfan Sinclair**, McGill University, Canada (Consultant on emulations and workflow models)
>
> **Melissa Terras**, Edinburgh University (Expert Adviser)

The **Development Team** included the following:

> **Spenser Mason** (Software Developer, USF AVC, Unity Programmer, Web Developer, WebGL Programmer)
>
> **Gilberto Jaimes** (USF AVC, 3D Modeling)
>
> **Marilyn Salvatierra** (USF AVC, Unity Programmer)
>
> **Lynette Kuliyeva** (USF English Department and Digital Humanities)

Additionally, the architect **Laura Romanò** helped us locate local records and photos of the site at via G. Ferraris 2, Gallarate, and shared materials on the recent renovation of the remaining "courtyard" building. Romanò, former staff members who worked with Busa, **Gisa Crosta** and **Livia Canestraro**, along with **Giovanni Bonetti** (of the Casa Di Francesco, the current occupant of the remaining building at the site), and **Filippo Proto**, all participated with Jones and Passarotti in a meeting and tour of the courtyard building and grounds on November 3, 2017. The meeting was coordinated and attended by Father Busa's former secretary, **Danila Cairati** (working locally with Passarotti).

Laura Romanò, Gisa Crosta, Marco Passarotti, Steven Jones, Danila Cairati, Filippo Proto, Livia Canestraro, via G. Ferraris, 2, Gallarate, November 3, 2017.

PRODUCTS AND ACCOMPLISHMENTS

PROJECT WEBSITE: http://www.recaal.org

The products available on the website are the result of convergent practices and methods, representing diverse ways of exploring the multiple dimensions of CAAL. So far, products include:

1. **Digitized materials from the Busa Archive** at the Università Cattolica del Sacro Cuore, Milan. For our project website, we've selected materials that are directly relevant to the establishment of CAAL, including documents, miscellaneous objects, and a small subset of the photographs commissioned and captioned by Father Busa himself. The materials in the Archive have only been accessioned and made available since Busa's death in 2011 and most remain unknown to the larger scholarly community. Our ultimate goal was to help jumpstart the process of digitizing the Busa Archive as a whole, a process now successfully underway under the direction of Senna and Passarotti at the Università Cattolica del Sacro Cuore in Milan.

2. **Oral-history interviews** with former women who worked at CAAL in the 1960s. Sample audio recordings with transcripts in Italian and English translations, by Nyhan and Passarotti--two so far--are included at the project website. These are part of a larger body of such interviews with CAAL's female punched-card machine operators, work already underway by Nyhan and Terras when they joined our project. Additional interviews are being published elsewhere by them. Among their other historical benefits, these stories complicate and challenge the origin myth surrounding Father Busa and the work of CAAL, while also contributing to the history of gender and technology.

3. **A multimodal essay reconstructing the workflow of CAAL**, produced by Rockwell and Sinclair, with development work by Andrew Macdonald, using some code from OpenSeadragon (https://openseadragon.github.io) and Storiiies (https://storiiies.cogapp.com/#storiiies).This dynamic interactive essay maps out what may have been key steps in the process (at least the idealized process) of "literary data processing," as designed at the beginning of Father Busa's work with IBM, ca. 1950-1952. The essay uses an IBM flowchart we digitized and have made available on the project website (in the Busa Archive section), and is linked to a punched-card emulator that uses a combination of Javascript, HTML, and image files, as well as an IPython program, to emulate one possible processes of entering data on the cards and tokenizing the linguistic data in the *Index Thomisticus*. This, too, is a kind of model, and it is based on Father Busa's and Paul Tasman's 1952 flowchart (which only came to the Archive in 2015). In other words, it's a model based on a prior model constructed by Busa himself.

4. **3D models of individual IBM machines** used at CAAL: the IBM 024/026 Card Punch, the 403 Alphabetical Accounting Machine, and the 082 Sorter. The work of CAAL was accomplished with a suite of machines, modular components in a data processing system, each component of which must be understood on its own and in relation to the system as a whole. The strangeness of these machines from the perspective of our own era, vividly captured in the manipulable and annotated 3D models, serves as a salutary reminder that the history of technology is not linear or progressive in any simple way. At the same time that CAAL was using machinery from IBM Italia in Milan, made available by negotiated agreements with IBM, elsdewehre, large-scale room-sized proto-mainframe computing systems were beginning to be used in industry and government. Indeed, some stages in CAAL's processing were performed on such systems, after transferring the data from punched cards to magnetic tape. Each individual 3D model is also integrated in the larger immersive model of CAAL as a whole, so that it can be understood as a component in a larger system.

5. **An immersive 3D "walkthrough" of CAAL.** This is the project's anchor model, built in Autodesk's Maya, then developed in the Unity Game Engine with WebGL to make it available online. It was produced by Jones, Kaplan, Mason, and others at USF's Advanced Visualization Center, under the direction of Kaplan. Besides a local version and the version for the web, another version was made in full Virtual Reality for the Oculus Rift. This has been demonstrated with USF graduate classes and visitors on a number of occasions in the AVC, using its large visualization wall while simultaneously connected to a user wearing a VR headset, making possible comparisons of VR, browser-based, and visualization-wall experiences of the modeled space.

Building the model posed special difficulties, even though CAAL was in operation only 50 years ago. Very little was known about this historically significant site before Jones's 2016 book and the RECAAL project (with the various independent research programs, already underway, that it gathered). We had to begin with the basics. Our November 3, 2017 meeting at via G. Ferraris 2, Gallarate (mentioned above) helped to establish the provisional dimensions and orientation of the now-demolished building that had once been attached to the existing square courtyard building, a long, former textile-factory, the interior of which appears in many Busa Archive photos and had housed CAAL's main operation. Previously, in November 2016, Danila Cairati had accompanied Jones to the site (as well as to other relevant locations in Gallarate), where they confirmed that the building was gone, having been demolished and replaced with a condominium building. Jones used Google Earth Pro's timeline feature to reveal aerial images of the demolished building dated March 2010:
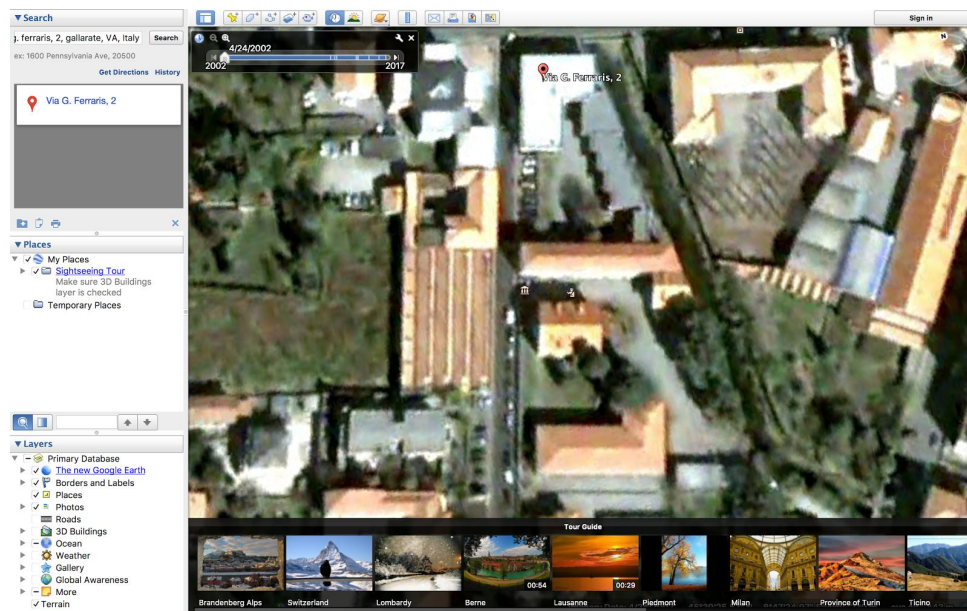
Google Earth Pro: Via G. Ferraris, 2, Gallarate. (Note the rectangular yellow building left of center, with what appear to be ridges on its roof, attached to the square courtyard building. Only the courtyard building now remains.)

A year after that first visit, on November 3, 2017, Jones and Passarotti returned to the site, this time met there by an expert in historical restoration, architect Laura Romanò, as well as two former CAAL workers and others. We were able to see the multiply-peaked "shadow" of the demolished building's skylight roof, faintly visible in outline against the south wall of the remaining courtyard building. And we measured the lot where the long building had been. Soon after the meeting, Romanò's searches turned up among local civic records some photos of the iconic long building before its demolition:

Via G. Ferraris, 2, Gallarate, sometime prior to 2012. (The taller building on the right is the square courtyard building that still stands. The long skylights on the roof and the windows along the wall match interior views in the Busa Archive photographs.)

One of the former workers brought to the meeting a memorial reconstruction, a hand-drawn, labeled sketch of CAAL's floorplan. This helped us to refine our understanding of the relationship between the existing site and the interiors of the demolished building and the equipment layout of CAAL, going well beyond the Busa Archive photos. Back at USF, all of these sources provided crucial information for the necessarily speculative work of building the immersive 3D model of CAAL's main building (the "walkthrough"). As a team, we understand these models to be heuristic aids to historical understanding, imaginative reconstructions, models in progress for thinking through the history of CAAL and the early humanities computing work done there.

At USF alone, a wide variety of scanning, modeling, and visualization projects have provided inspiration and collaborative instruction for us. One group (the Institute for Digital Exploration, IDEx) used LIDAR scanning to model archaeological sites in Sicily in pursuit of better historical understanding. Another group, Access3D, in collaboration with the Advanced Visualization Center (AVC), used drones, photogrammetry, and LIDAR to model the lighthouse and surroundings on Egmont Key, Florida. They're also working with the Seminole people to reconstruct a little-known cultural-heritage layer at the same site: the island was used as an internment camp for Seminole people during the Indian Removal Period, 1856-1858. Another group in the USF Library's Digital Heritage and Humanities Collections, using LIDAR and software modeling, is capturing the infrastructural remains of the mid-century US space program at Cape Canaveral, from building complexes to launch pads.

The USF Advanced Visualization Center itself has had a good deal of experience with relevant projects, including a 3D historical simulation of the fire at the Dozier School for Boys on November 18, 1914, and the Virtual Patient Safety project, a VR environment for training nursing staff on proper procedures for preventing and responding to patient falls.

RECAAL has learned from all of these examples. We are focused on the history of mid-century technology infrastructure, but also on the nearly-lost infrastructure of practice: workflow, invisible cultural layers, inaccessible human actions in the historical environment, and an all-but-forgotten program of work in the space, all in the larger contexts of the cold war, and the specific technical processes and machineries that already run the risk of becoming obscured by time.

Screenshot from RECAAL's 3D model ("walkthrough") of CAAL, looking south from the entrance.

LONG-TERM IMPACT, ONGOING QUESTIONS

Our collaborative reconstruction project has contributed to the history of early humanities computing and Roberto Busa's role in it. Already the founding myth has been complicated with the particulars we've recovered, and we expect this trend to continue with additional modeling, the increased availability of materials in the Busa Archive, and the appearance of additional publications and presentations based on these materials and our models. The following are some of the specific

outcomes--beyond the website and its assets as they are now--and our additional plans for the project going forward, beyond the funded period.

1. Both Jones and Nyhan are scheduled to deliver keynote lectures about the project at the AIUCD Digital Humanities conference in Milan, January 15-17, 2020.

2. A curator at the National Museum of American History, the Smithsonian Institution in Washington, D.C., has discussed with Jones the possibility of including material from the project in a future exhibit on technology and religion in American history. (As part of that plan, see #5 below.)

3. The website for the project will be hosted by Research Computing and the AVC at USF and will continue to develop there. It will be mirrored at McGill University in Canada, beginning early in 2020. All materials produced by the project are available open access, free of charge but with some variant Creative Commons licenses applied, depending on their institutional origins. All relevant source code is available via GitHub or upon request.

4. The larger contents of the Busa Archive, Università Cattolica del Sacro Cuore, Milan, will continue to be digitized and made available at the Archive's own website, under the direction of Senna and Library staff, in consultation with Passarotti.

5. We plan to continue to refine and iterate our 3D models, further integrating newly digitized materials in the Busa Archive with the "walkthrough," for example, and animating objects inside the model (punched-card drawers or the consoles of machiners), perhaps adding sounds to better represent the ambience of the punched-card machinery line during operation. A VR version of the "walkthrough" is now being developed for the Oculus Quest, a self-contained, all-in-one headset, allowing us to carry an immersive version of the model of CAAL with us to classrooms, conferences, meetings, and exhibits.

NEWLY RAISED OR ONGOING RESEARCH QUESTIONS

- What machines, precisely, were used at various stages during the six-year period? The photographs are inevitably limited and staged, and they show only what was set up on the day of photography. (Indeed, we have seen some personal snapshots of the former punched-card operators that suggest the day-to-day layout of CAAL was less formal and regimented than *any* of the curated images in the Busa Archive would suggest. For example, they show many workstations with *two* female punched-card operators working side by side, unlike in the Archive images.)

- What exactly are the operators (often in pairs) doing at the tables to the side of the machine-lines in the photos? Presumably they're engaged in some form of sorting, but, based on interviews, it's unlikely to be higher-level linguistic work such as lemmatization, for example.

- Were the plug boards we can see displayed in photos on top of the cabinets functional or merely for demonstration? If the former, who programmed them?

- What was the story behind the encoding protocols used for the punched cards at CAAL? How does it map onto the development of encoding standards such as BCDIC, EBCDIC, etc.?

- Was there lab "staff" besides the operators, as the "front office" space apparent in the photos and layout of CAAL would seem to imply, and what were their roles?

- How precisely did work done at CAAL contribute to the development inside IBM of KWIC (Key Word In Context) and other protocols for Information Retrieval and emerging Natural Language Processing (as Paul Tasman has suggested it did)?

- How, specifically, did Busa coordinate work at CAAL with emergent "centers" in other locations (as was his expressed interest), in Nancy, France, for example, or in Germany, or at EURATOM, Ispra, Italy?

- What was the terminus of CAAL's work on the Dead Sea Scrolls? Was it 1959, as the correspondence would suggest? Or was any further work on that project conducted at CAAL in the 1960s?

RESULTANT PUBLICATIONS AND PUBLIC PRESENTATIONS

Jones, Steven E., Julianne Nyhan, Geoffrey Rockwell, Marco Passarotti. Symposium in DHLabs, USF, Tampa, Florida, March 19, 2019, open to the public: "Reconstructing Roberto Busa's CAAL." (This represented a collaborative summative consideration of the work of the project, followed by audience discussion and a demonstration of the 3D "walkthrough.")

Jones, Steven E. "Reverse Engineering the First Humanities Computing Center." *Digital Humanities Quarterly* 12.2 (2018): http://www.digitalhumanities.org/dhq/vol/12/2/000380/000380.html.

Jones, Steven E. Invited lecture: "'A New Humanism': Expo '58, Roberto Busa, and the First Humanities Computing Center," Centre for Contemporary and Digital History (C²DH), The University of Luxembourg (November 6, 2019):

https://www.c2dh.uni.lu/data/lecture-steven-e-jones-new-humanism-expo-58-roberto-busa-and-first-humanities-computing-center.

Jones, Steven E. Invited lecture: "The Cold War and a 'New Humanism': The History of the First Humanities Computing Center," New College of Florida (September 26, 2019).

Jones, Steven E. Invited presentation: "Roberto Busa, S.J.'s Literary Data Processing," Religion and Innovation Symposium, Smithsonian National Museum of American History, Washington, D.C. (April 12, 2019).

Jones, Steven E. "N-Dimensional Modeling of the First Humanities Computing Center," on panel: "DH in 3D: Multidimensional Research and Education in the Digital Humanities," ADHO Digital Humanities Conference, 2018, Mexico City (June 27, 2018).

Jones, Steven E. Invited keynote lecture: "Reconstructing the First Humanities Computing Center: Modeling What we Don't know," Digital Humanities Showcase (Vitrine Humanités Numériques), University of Montreal (January 26, 2018).

Jones, Steven E., with coauthors: Julianne Nyhan, Geoffrey Rockwell, Stefan Sinclair, Melissa Terras, "Reverse Engineering the First Humanities Computing Center." ADHO Digital Humanities Conference, 2017, Montreal (August 9, 2017).

Nyhan, Julianne and Marco Passarotti, ed. *One Origin of Digital Humanities: Father Roberto Busa in His Own Words.* Springer, 2019. (Foreword by Steven E. Jones.)

Rockwell Geoffrey, and Marco Passarotti. "The Index Thomisticus as a Big Data Project." *Umanistica Digitale* 5 (2019): 13-34: DOI: http://doi.org/10.6092/issn.2532-8816/8575.

Terras, Melissa and Julianne Nyhan. "Father Busa's Female Punched-Card Operators." In *Debates in Digital Humanities 2016.* Eds. Matthew K. Gold and Lauren Klein. Minneapolis: University of Minnesota Press: 2016: http://dhdebates.gc.cuny.edu/debates/text/57.